

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

JC921 U.S. PTO
09/708575
11/09/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application:

1999年11月 9日

出願番号
Application Number:

平成11年特許願第318819号

出願人
Applicant(s):

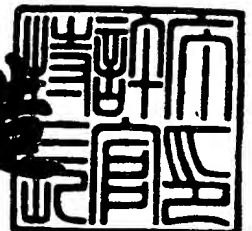
株式会社東芝

CERTIFIED COPY OF
PRIORITY DOCUMENT

2000年 9月 1日

特許庁長官
Commissioner,
Patent Office

及川耕造



【書類名】 特許願

【整理番号】 A009906028

【提出日】 平成11年11月 9日

【あて先】 特許庁長官 殿

【国際特許分類】 G06K 9/00

【発明の名称】 所在情報認識方法と認識アプリケーションと記録媒体と
所在情報認識装置

【請求項の数】 10

【発明者】

 【住所又は居所】 神奈川県川崎市幸区柳町 7 0 番地 株式会社東芝柳町工
 場内

 【氏名】 中尾 昭彦

【特許出願人】

 【識別番号】 000003078

 【氏名又は名称】 株式会社 東芝

【代理人】

 【識別番号】 100058479

 【弁理士】

 【氏名又は名称】 鈴江 武彦

 【電話番号】 03-3502-3181

【選任した代理人】

 【識別番号】 100084618

 【弁理士】

 【氏名又は名称】 村松 貞男

【選任した代理人】

 【識別番号】 100068814

 【弁理士】

 【氏名又は名称】 坪井 淳

【選任した代理人】

【識別番号】 100092196

【弁理士】

【氏名又は名称】 橋本 良郎

【選任した代理人】

【識別番号】 100091351

【弁理士】

【氏名又は名称】 河野 哲

【選任した代理人】

【識別番号】 100088683

【弁理士】

【氏名又は名称】 中村 誠

【選任した代理人】

【識別番号】 100070437

【弁理士】

【氏名又は名称】 河井 将次

【手数料の表示】

【予納台帳番号】 011567

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 所在情報認識方法と認識アプリケーションと記録媒体と所在情報認識装置

【特許請求の範囲】

【請求項 1】 紙葉類上に記載され、種々の国ごとに異なる複数段の階層構造のカテゴリーにより構成されている所在情報を認識するものにおいて、

種々の国ごとに対応して設けられ、上記所在情報を認識するための複数の辞書と、

上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリーに対する種々の認識手順とを有し、

上記辞書を選択し、

上記認識手順を選択する選択し、

上記紙葉類上に記載されている所在情報を読み取り、

この読み取った所在情報を上記選択されている認識手順にそって、かつ上記選択されている辞書を用いて認識する、

ことを特徴とする所在情報認識方法。

【請求項 2】 種々の国ごとに異なる複数段の階層構造のカテゴリーにより構成されている所在情報を認識する認識アプリケーションにおいて、

種々の国ごとに対応して設けられ、上記所在情報を認識するための複数の辞書と、

上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリーに対する種々の認識手順とを有し、

上記所在情報を認識する際に、上記辞書の 1 つが選択され、上記認識手順の 1 つが選択され、この選択された辞書と認識手順に基づいて、認識処理が行われることを特徴とする認識アプリケーション。

【請求項 3】 種々の国ごとに異なる複数段の階層構造のカテゴリーにより構成されている所在情報を認識するものに用いられる記録媒体において、

種々の国ごとに対応し、上記所在情報を認識するための複数の辞書と、

上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリーに対

する種々の認識手順とが記録されていることを特徴とする記録媒体。

【請求項 4】 所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と

、
上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

この所在情報単語検出手段で得られた第 1 の単語領域内に含まれる文字情報を認識対象である地域内に存在する地名を登録した単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第 1 の単語認識手段と、

この第 1 の単語認識手段で処理した第 1 の単語領域と、この第 1 の単語領域と同一行内で隣接する第 2 の単語領域とを結合した第 3 の単語領域内に含まれる文字情報を上記単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第 2 の単語認識手段と、

上記第 1 の単語認識手段による認識結果の単語評価値と上記第 2 の単語認識手段による認識結果の単語評価値を比較し、単語評価値の高い方の認識結果を出力する出力手段と、

を具備することを特徴とする所在情報認識装置。

【請求項 5】 所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と

、
上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

この所在情報単語検出手段で得られた第 1 の単語領域内に含まれる文字情報を認識対象である地域内に存在する地名を登録した単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第 1 の単語認識手段と、

この第 1 の単語認識手段で処理した第 1 の単語領域内に含まれる文字情報が第 1 の単語領域を複数の単語に分割する条件を満たしているか否かを判断する判断手段と、

この判断手段により複数の単語に分割する条件を満たしていると判断された際に、上記分割された各単語を上記単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第 2 の単語認識手段と、

上記第 1 の単語認識手段による認識結果の単語評価値と上記第 2 の単語認識手段による各単語の認識結果の単語評価値を比較し、単語評価値の高い方の認識結果を出力する出力手段と、

を具備することを特徴とする所在情報認識装置。

【請求項 6】 上記判断手段により文字情報を複数の単語に分割する条件が、単語を構成する所定の文字間の距離が同一単語内の他の文字間の距離と比べて大きい場合に、満足するものであることを特徴とする請求項 5 に記載の所在情報認識装置。

【請求項 7】 複数段の階層構造のカテゴリにより構成されている所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、

上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも 1 つの文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

上記所在情報を構成する複数段の階層構造の各カテゴリに対応する上記所在情報単語検出手段で得られた各単語領域での単語を認識する順序を設定する設定

手段と、

この設定手段により設定される各単語領域に対する認識の順序にしたがって、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の 1 つの単語辞書の内容と照合することにより単語を認識する単語認識手段と、

この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段と、

を具備することを特徴とする所在情報認識装置。

【請求項 8】 複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と

上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも 1 つの文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

上記所在情報を構成する複数段の階層構造の各カテゴリーに対応する上記所在情報単語検出手段で得られた各単語領域での単語を認識する順序があらかじめ記憶されている IC と、

この IC に記憶されている各単語領域に対する認識の順序にしたがって、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の 1 つの単語辞書の内容と照合することにより単語を認識する単語認識手段と

この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段と、

を具備することを特徴とする所在情報認識装置。

【請求項 9】 複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、

上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも 1 つの文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の 1 つに対応し、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報で構成する複数の文字列の組み合わせの少なくとも一部が一致する上記単語辞書内の 1 個または複数個の単語を抽出する単語抽出手段と、

上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、上記単語抽出手段により抽出した 1 個または複数個の単語と照合することにより単語を認識する単語認識手段と、

この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段と、

を具備することを特徴とする所在情報認識装置。

【請求項 10】 複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、

この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、

上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、

上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも 1 つの文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、

この所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識

対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の 1 つの単語辞書の内容と照合することにより単語を認識する単語認識手段と、

この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段と、

を具備し、

上記単語認識手段が、

所定のカテゴリーに対応する上記単語辞書の登録単語数が所定数以上の際に、上記文字情報を構成する複数の文字列の組み合わせの少なくとも一部が一致する上記単語辞書内の単語を 1 個または複数個抽出する単語抽出手段と、

この単語抽出手段により抽出した 1 個または複数個の単語と上記文字情報を照合することにより単語を認識する第 1 の認識手段と、

所定のカテゴリーに対応する上記単語辞書の登録単語数が所定数未満の際に、上記単語辞書の内容と上記文字情報を照合することにより単語を認識する第 2 の認識手段とからなる、

ことを特徴とする所在情報認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、所在情報としての住所を認識する所在情報認識方法と認識アプリケーションと記録媒体と所在情報認識装置に関する。

【0002】

【従来の技術】

一般に、葉書や名刺などに書かれている住所情報（所在情報）を光学式文字読取装置（OCR 装置）によって光学的に読み取る場合、まずその書状の画像を取りこんでから住所の記載されている領域を指定または推定し、その領域内から行や文字を切り出す。

【0003】

OCR 装置内には認識対象地域内の地名辞書が用意しており、住所領域内に書

かれている文字をこの辞書と照らし合わせながら読んでいくことによって住所認識を行う。

【0004】

住所認識の方式としては、日本の場合だとまず都道府県名や市名といったおおまかな地域情報の文字列を検出し、その続きの文字列を町名などのより細かい地域情報として読んでいく方法が一般的である。この後、特定の文字や文字列を検出するなど、住所認識率を上げるためにさまざまな工夫が考案されている。

【0005】

以下では、探索パターン列が文字認識処理によって得られた文字列、辞書パターン列が単語辞書に登録されている住所名文字列候補である場合について、具体的に述べる。

【0006】

まず、装置の汎用性について説明する。

【0007】

たとえば、国が異なると住所の記載書式は全く違っていることが多い。例えば日本では大まかな地域名から順番に書くのが普通であるが、欧米では細かい地域情報から順に書くことが多く、最初にストリート名を書き、その後に都市名や州名を書く。そのため国が変わると住所認識を行うための地名辞書が変わるだけでなく、住所認識の手順も変える必要がある。

【0008】

国による住所認識手順の違いは、汎用的な住所認識装置を開発する際に大きな問題となる。例えば英語圏用に開発した住所認識装置でフランス語圏の住所を認識しようとして、地名辞書だけをフランス語圏用に修正しても十分な性能を得られない。フランス語圏用の住所認識手順を導入する必要があるが、装置の回路を各国用に調整することはコスト高の要因となる。

【0009】

次に、類似地名の誤認識について説明する。

【0010】

たとえば、ある地域に「YORK」「NORTH YORK」「EAST Y

OPK」いう都市名が存在した場合を考える。その地域の住所を認識する際に、住所行の一部が「YORK」と認識できたとしても、実際にそこに書かれている都市名は「NORTH YORK」かも知れない。

【0011】

逆に「EAST YORK」と認識した場合でも、「EAST」の部分は別の単語を誤認識している可能性がある。

【0012】

次に、単語絞り込み辞書サイズの肥大化について説明する。

【0013】

たとえば、1つの国の国内住所を全て認識できるようにするには、当然国内の全地名を住所認識用の単語辞書に登録する必要がある。しかし高速に住所認識を行うにはさらに単語辞書に情報を追加する必要がある。

【0014】

例えば「ABC」という大都市に1000以上のストリートが存在したとする。この場合、ABCという都市のストリート名を認識するためには、ストリート名の探索パターン列の位置が分かっているとしても1000回以上、辞書パターン列との比較処理を実行する必要となる。

【0015】

比較回数を減らす方法の1つとして、探索パターン列の特徴から比較対象とする辞書パターン列をある程度絞り込み、絞り込んだ辞書パターン列と探索パターン列を比較する方法がある。

【0016】

探索パターンがアルファベットなど文字種数が少ない場合によく用いられるのがbigram(N-gramという手法で、N=2とした場合を示している)と呼ばれる手法である。これはAB、BC、…、ZZといった2文字の並びのそれぞれについて、その2文字の並びを含む辞書パターン列のリストをあらかじめ作成しておく方法である。

【0017】

このbigramは、

- ・ 文字種数が少なく、
- ・ 文字間にゴミが入りやすい

場合の文字認識に有効である。

【0018】

例えば「JOHNSON」という辞書パターン列は「JO」「OH」「HN」「NS」「SO」「ON」のリストに登録される全ての2文字の並びについて辞書パターン列のリストを作成したもののことを以下では単語絞り込み辞書と呼ぶことにする。

【0019】

探索パターン列と単語辞書に登録された辞書パターン列の比較を実行する前に、探索パターン列に含まれる2文字の並びを調べ、それらを含む辞書パターン列に得点を与えていく。そして総得点の高い辞書パターン列を選択し、それらと探索パターン列とを比較することによって単語認識を行う。例えば総得点上位10位までを使用することにとすると、ストリート数が1000以上ある都市のストリート名を認識する場合だと探索パターン列と辞書パターン列の比較処理の回数は1/100以下となる。

【0020】

ただし、認識対象地域内の全ての都市やストリート名について単語絞り込み辞書を用意した場合、単語辞書の総容量より単語絞り込み辞書の総容量の方がはるかに多くなることが多い。

【0021】

【発明が解決しようとする課題】

この発明は、上記欠点を除去するもので、わずかな修正を加えるだけで各国の所在情報認識を実行できる所在情報認識方法と認識アプリケーションと記録媒体と所在情報認識装置を提供することを目的としている。

【0022】

【課題を解決するための手段】

この発明の所在情報認識方法は、紙葉類上に記載され、種々の国ごとに異なる複数段の階層構造のカテゴリーにより構成されている所在情報を認識するものに

において、種々の国ごとに対応して設けられ、上記所在情報を認識するための複数の辞書と、上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリに対する種々の認識手順とを有し、上記辞書を選択し、上記認識手順を選択する選択し、上記紙葉類上に記載されている所在情報を読み取り、この読み取った所在情報を上記選択されている認識手順にそって、かつ上記選択されている辞書を用いて認識するものである。

【 0 0 2 3 】

この発明の認識アプリケーションは、種々の国ごとに異なる複数段の階層構造のカテゴリにより構成されている所在情報を認識するものにおいて、種々の国ごとに対応して設けられ、上記所在情報を認識するための複数の辞書と、上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリに対する種々の認識手順とを有し、上記所在情報を認識する際に、上記辞書の 1 つが選択され、上記認識手順の 1 つが選択され、この選択された辞書と認識手順に基づいて、認識処理が行われるものである。

【 0 0 2 4 】

この発明の記録媒体は、種々の国ごとに異なる複数段の階層構造のカテゴリにより構成されている所在情報を認識するものに用いられるものにおいて、種々の国ごとに対応し、上記所在情報を認識するための複数の辞書と、上記種々の国ごとに異なり、所在情報の複数段の階層構造の各カテゴリに対する種々の認識手順とが記録されているものである。

【 0 0 2 5 】

この発明の所在情報認識装置は、所在情報画像を読み取る読取手段と、この読取手段により読み取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読み取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる文字行を 1 つまたは複数の単語領域に分割する所在情報単語検出手段と、この所在情報単語検出手段で得られた第 1 の単語領域内に含まれる文字情報を認識対象である地域内に存在する地名を登録した単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値

を出力する第1の単語認識手段と、この第1の単語認識手段で処理した第1の単語領域と、この第1の単語領域と同一行内で隣接する第2の単語領域とを結合した第3の単語領域内に含まれる文字情報を上記単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第2の単語認識手段と、上記第1の単語認識手段による認識結果の単語評価値と上記第2の単語認識手段による認識結果の単語評価値を比較し、単語評価値の高い方の認識結果を出力する出力手段とからなる。

【0026】

この発明の所在情報認識装置は、所在情報画像を読取る読取手段と、この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる文字行を1つまたは複数の単語領域に分割する所在情報単語検出手段と、この所在情報単語検出手段で得られた第1の単語領域内に含まれる文字情報を認識対象である地域内に存在する地名を登録した単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第1の単語認識手段と、この第1の単語認識手段で処理した第1の単語領域内に含まれる文字情報が第1の単語領域を複数の単語に分割する条件を満たしているか否かを判断する判断手段と、この判断手段により複数の単語に分割する条件を満たしていると判断された際に、上記分割された各単語を上記単語辞書の内容と照合することにより単語を認識するとともに、認識結果の単語評価値を出力する第2の単語認識手段と、上記第1の単語認識手段による認識結果の単語評価値と上記第2の単語認識手段による各単語の認識結果の単語評価値を比較し、単語評価値の高い方の認識結果を出力する出力手段とからなる。

【0027】

この発明の所在情報認識装置は、複数段の階層構造のカテゴリにより構成されている所在情報画像を読取る読取手段と、この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検

出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも1つの文字行を1つまたは複数の単語領域に分割する所在情報単語検出手段と、上記所在情報を構成する複数段の階層構造の各カテゴリーに対応する上記所在情報単語検出手段で得られた各単語領域での単語を認識する順序を設定する設定手段と、この設定手段により設定される各単語領域に対する認識の順序にしたがって、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の1つの単語辞書の内容と照合することにより単語を認識する単語認識手段と、この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段とからなる。

【 0 0 2 8 】

この発明の所在情報認識装置は、複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも1つの文字行を1つまたは複数の単語領域に分割する所在情報単語検出手段と、上記所在情報を構成する複数段の階層構造の各カテゴリーに対応する上記所在情報単語検出手段で得られた各単語領域での単語を認識する順序があらかじめ記憶されているICと、このICに記憶されている各単語領域に対する認識の順序にしたがって、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の1つの単語辞書の内容と照合することにより単語を認識する単語認識手段と、この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段とからなる。

【 0 0 2 9 】

この発明の所在情報認識装置は、複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読取った所在情

報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも1つの文字行を1つまたは複数の単語領域に分割する所在情報単語検出手段と、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の1つに対応し、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報で構成する複数の文字列の組み合わせの少なくとも一部が一致する上記単語辞書内の1個または複数個の単語を抽出する単語抽出手段と、上記所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、上記単語抽出手段により抽出した1個または複数個の単語と照合することにより単語を認識する単語認識手段と、この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段とからなる。

【0030】

この発明の所在情報認識装置は、複数段の階層構造のカテゴリーにより構成されている所在情報画像を読取る読取手段と、この読取手段により読取った所在情報画像から文字行を検出する行検出手段と、上記読取手段により読取った所在情報画像から所在情報が記載されている領域を検出する領域検出手段と、上記行検出手段で検出した文字行のうち上記領域検出手段で検出した所在情報領域内に含まれる少なくとも1つの文字行を1つまたは複数の単語領域に分割する所在情報単語検出手段と、この所在情報単語検出手段で得られた単語領域内に含まれる文字情報を、認識対象である地域内に存在する上記カテゴリーごとに異なる地名を登録した複数の単語辞書の1つの単語辞書の内容と照合することにより単語を認識する単語認識手段と、この単語認識手段による各カテゴリーに対応する認識結果を上記所在情報の認識結果として出力する出力手段とを具備し、上記単語認識手段が、所定のカテゴリーに対応する上記単語辞書の登録単語数が所定数以上の際に、上記文字情報を構成する複数の文字列の組み合わせの少なくとも一部が一致する上記単語辞書内の単語を1個または複数個抽出する単語抽出手段と、この単語抽出手段により抽出した1個または複数個の単語と上記文字情報を照合することにより単語を認識する第1の認識手段と、所定のカテゴリーに対応する上記

単語辞書の登録単語数が所定数未満の際に、上記単語辞書の内容と上記文字情報を照合することにより単語を認識する第 2 の認識手段とからなる。

【0031】

【発明の実施の形態】

以下、図面を参照してこの発明の実施形態を説明する。

【0032】

すなわち、まずわずかな修正を加えるだけで各国の住所認識（所在情報認識）を実行できる汎用的な住所認識装置（所在情報認識装置）の一例について説明する。

【0033】

図 1 は、この発明の住所認識装置の概略構成を示している。

【0034】

上記住所認識装置は、所在情報としての住所情報が記載されている郵便物等の書状（紙葉類）S から表面の画像を光電変換によって取り込む（読取る）画像取り込み部（読取手段）1、この画像取り込み部 1 により取り込んだ画像により住所の記載されている領域を検出する領域検出部 2、この領域検出部 2 により検出した住所の記載領域から住所の単語を検出する住所単語検出部 3、この住所単語検出部 3 からの住所の単語と住所辞書 4 に記憶されている住所との比較により単語の認識処理する単語認識処理部 5、この単語認識処理部 5 における認識処理の手順と使用する住所辞書 4 が設定されている住所書式設定部 6、上記各部を制御する住所認識制御部 7、この住所認識制御部 7 に得られた住所認識の結果を出力する住所認識結果出力部 8 により構成されている。

【0035】

上記領域検出部 2 は、領域を 1 つだけ検出してもよいし、複数の領域を検出して可能性の高い順に処理してもよい。

【0036】

上記住所単語検出部 3 は、領域検出部 2 が検出した領域内から住所行を見つけ、さらに行から文字を切り出したり行から単語を切り出すなどの処理を行うものである。

【0037】

上記住所認識制御部7は、住所書式設定部6から与えられたルールに従って認識したい単語を順次、単語認識処理部5に送り、単語認識処理部5から返ってきた認識結果を見ながら、次に認識すべき単語を決めたり単語の読み直しを行ったりするものである。

【0038】

上記住所の記載方法としては、日本等において、郵便番号、都道府県名、市区名、町名、街区の順に、たとえば1番上の行から順にしかも左から右へ記載されるようになっている。住所の地域を示す階層構造の上位のカテゴリのもののから順に記載されるようになっている。

【0039】

これに対して、カナダ（欧米）等では、上記住所の記載方法として、1番下の行から順にしかも右側から順に郵便番号、州名、都市名、ストリート名、ストリート番号の順に記載されるようになっている。

【0040】

たとえば、図1に示すように、「123 ABC STREET TORONTO ONTARIO Z9Z 9Z9」となっている。

【0041】

上記住所書式設定部6により設定される認識処理の手順としては、認識対象としている国や地域の住所記載書式に関する情報や住所領域の検出するためのテクニック、または住所認識処理時のテクニックなどをルールとして設定するものである。この設定としては切替えスイッチ等のハードウェアで行う方法もあるし、設定ファイルを用意しておいてそれを装置が読み取る方式も考えられる。住所書式設定部6が読み込んだ情報は住所認識制御部7に送られる。

【0042】

このように、上記住所書式設定部6で与える情報を変えることにより、同じ住所認識装置で異なる国の住所を扱うことが可能となる。

【0043】

上記住所書式設定部6により設定される認識処理の手順としての日本用の住所

認識ルールの例について説明する。

【0044】

すなわち、

- ・ 単語は行の前から読んでいく
- ・ 単語をたどる順は行頭から行末へ
- ・ 一番最初に郵便番号を読む
- ・ 郵便番号の単語の続きから都道府県名单語を探す
- ・ 都道府県名单語の続きから市区名单語を探す
- ・ 市区名单語の続きから町名单語を探す
- ・ 町名单語の続きの単語を街区情報として認識する

また、上記住所書式設定部 6 により設定される認識処理の手順としてのカナダ用の住所認識ルールの例について説明する。

【0045】

すなわち、

- ・ 単語は行の後ろから読んでいく
- ・ 単語をたどる順は行末から行頭へ
- ・ 一番最初に郵便番号を読む
- ・ 郵便番号の単語の続きから州名单語を探す
- ・ 州名单語の続きから都市名单語を探す
- ・ 都市名单語の続きからストリート名单語を探す
- ・ ストリート名单語の続きの単語をストリート番号として認識する

上記住所書式設定部 6 の構成としては、図 2 のように、あらかじめ住所読み取りルールを記述したファイルを用意しておき、そのファイルを読取ることで住所認識装置に読み取りルールを教える方式がまず考えられる。この場合、住所書式設定部 6 は、住所認識ルールファイル 6 a とこれを読取る住所認識ファイル読取部 6 b により構成されている。

【0046】

しかしこの方式だと、

- ・ 工場から出荷する時に住所認識装置 1 台ごとに住所認識ルールファイルをロ

ードする手間が面倒。

【0047】

・ファイル情報のセキュリティが甘く、第3者が住所書式設定ルールを盗み出すことが容易。

【0048】

といった問題が発生する。

【0049】

各国用の住所単語辞書4は、引越し、家屋の新築、市区町村名の統廃合などの理由で頻繁に変更が必要である。しかし住所書式設定情報というものは、一度設定してしまうと大きな修正を加える必要はほとんどない。そこで図3のように住所書式設定ルールをICに焼き、そのICからルールを読み出す方式にしてもよい。この場合、住所書式設定部6は、住所認識ルールIC6cとこのIC6cの読取りを行う住所認識ファイルIC読取部6dにより構成されている。

【0050】

この際、ルールの解析は、ファイルで持つよりはるかに困難になるためセキュリティが高まる。またICを住所認識装置の住所認識ファイルIC読取部に刺す（装填する）だけで住所書式設定情報をロードすることが可能になる。また住所設定ルールを焼き込んだICを交換するだけで、各国の住所認識用に住所認識装置に設定できるようにしても良い。この場合、住所設定ルールと住所辞書とを国ごとにペアで交換することができる。

【0051】

上記住所辞書4としては、日本用の住所辞書4aと、カナダ用の住所辞書4bとが用意されている。

【0052】

日本用の住所辞書4aとしては、都道府県名の単語辞書、各都道府県ごとの市区名の単語辞書、各市区ごとの町名の単語辞書が用意されている。

【0053】

カナダ用の住所辞書4bとしては、図4から図6に示すように、州名の単語辞書11、各州ごとの都市名の単語辞書12、…、各都市ごとのストリート名の単語

語辞書 13、…が用意されている。

【0054】

上記したように、住所書式設定部により住所書式設定ルールと住所辞書を設定することができる。すなわち、所定の国に対応した、住所書式設定ルールと住所辞書を選択することができる。

【0055】

また、画像取り込み部 1、領域検出部 2、住所単語検出部 3、単語認識処理部 5、住所認識制御部 7、住所認識結果出力部 8 が、認識処理のアプリケーションと、住所書式設定部と住所辞書のアプリケーションとからなり、住所書式設定部で設定された住所書式設定ルールと住所辞書に基づいて認識処理のアプリケーションが認識処理を実行するようにしても良い。

【0056】

また、住所書式設定部と住所辞書とが CD、DVD 等の記録媒体に記録されており、上記画像取り込み部 1、領域検出部 2、住所単語検出部 3、単語認識処理部 5、住所認識制御部 7、住所認識結果出力部 8 からなる認識処理装置に、記録媒体再生部を設け、この記録媒体再生部により再生された住所書式設定部の内容に基づいて、住所書式設定ルールと住所辞書とを設定し、この設定された内容で認識処理装置が認識処理を実行するようにしても良い。

【0057】

次に、類似地名の誤認識の防止について説明する。

【0058】

ある地域に「YORK」「NORTH YORK」「EAST YORK」の 3 つの都市が存在した場合を考える。その地域の住所を認識する際に、住所行の一部が「YORK」と認識できたとしても、実際にそこに書かれている都市名は「NORTH YORK」かも知れない。

【0059】

「YORK」と「NORTH YORK」の両方を区別することが可能な住所単語認識処理の一例を図 7 のフローチャートに示す。基本的には住所認識制御部 7 から教えられた単語認識処理の開始位置から 1 単語ずつ住所単語辞書 4 を使っ

て認識していく。しかしこれだけでは「YORK」は読めても複数単語から成る「NORTH YORK」は読めないため、図 8 に示すように、現在処理中の単語（「YORK」）W1 と単語 W1 に隣接する単語（「NORTH」）W2 をくっつけて新しい単語（「NORTH YORK」）W3 を作成し、単語 W3 を認識してみる。図 7 では 2 単語を接続する場合のみを取り上げているが、3 単語以上を接続する場合もありうる。

【0060】

そして、1 単語だけで単語認識した場合と、複数単語をくっつけて作成した単語を単語認識にかけた場合を比較し、結果の良かった方を採用する。認識結果の評価値があらかじめ設定してあるしきい値より低い場合はどちらの単語認識結果も採用せず、単語 W1 の次に書かれている単語を新たな単語 W1 として上記の処理を繰り返す。

【0061】

上記住所認識制御部 7 による住所単語認識処理について、図 7 に示すフローチャートを参照しつつ説明する。

【0062】

すなわち、住所認識制御部 7 は、住所単語認識処理を開始し、住所単語探索開始位置へ移動する（ST1）。たとえば、カナダ用の住所認識方法に設定されている場合、最終行の後ろから順に読んでいく。

【0063】

この際、住所認識制御部 7 は、認識処理にかけていない単語が存在しなかった場合（ST2）、単語認識エラー処理へ移行する。

【0064】

上記住所認識制御部 7 は、ステップ 2 により認識処理にかけていない単語が存在した場合、単語を 1 つ選択し、選択した単語 W1 を与えられた地名辞書（11、12、13）を用いて単語認識処理する（ST3）。たとえば、選択した単語 W1 が州名に対応する単語の場合、単語辞書 11 を用い、選択した単語 W1 が都市名に対応する単語の場合、上記州名に対応する単語辞書 12 を用い、選択した単語 W1 がストリート名に対応する単語の場合、上記都市名に対応する単語辞書

13を用いる。

【0065】

この結果、住所認識制御部7は、単語認識結果A1、単語評価値S1とを算出する(ST3)。

【0066】

次に、住所認識制御部7は、単語W1の続きの位置にまだ認識処理にかけていない単語W2が存在するか否かを判断する(ST4)。

【0067】

住所認識制御部7は、単語W2が存在すると判断した場合、単語W1と単語W2を接続して新しい単語W3を作成し(ST5)、この作成した単語W3を対応する地名辞書(11、12、13)を用いて単語認識処理する(ST6)。

【0068】

この結果、住所認識制御部7は、単語認識結果A3、単語評価値S3とを算出する(ST6)。

【0069】

これにより、住所認識制御部7は、単語W1に対する一番高い単語評価値S1と、単語W3に対する単語評価値S3が一番高い単語評価値S3とを比較し、単語W3に対する一番大きな単語評価値S3が単語W1に対する一番大きな単語評価値S1よりも大きいか同じで、かつ単語W3に対する一番大きな単語評価値S3が所定のしきい値以上の際に(ST7)、単語W3に対する単語認識結果A3を認識結果として出力する。

【0070】

また、上記住所認識制御部7は、上記比較により単語W1に対する一番大きな単語評価値S1が単語W3に対する一番大きな単語評価値S3よりも大きく、かつ単語W1に対する一番大きな単語評価値S1が所定のしきい値以上の際に(ST8)、単語W1に対する単語認識結果A1を認識結果として出力する。

【0071】

また、上記住所認識制御部7は、上記ステップ7、8を満足しなかった場合、ステップ2に戻る。

【0072】

また、上記住所認識制御部7は、上記ステップ4において、単語W2が存在しないと判断した場合、単語W3に対する単語評価値S3を「0」とし(ST9)、ステップ7に進む。

【0073】

この場合の例を図8を用いて説明する。

【0074】

すなわち、都市名の単語(「YORK」)W1と単語W1に隣接する単語(「NORTH」)W2をくっつけて新しい単語(「NORTH YORK」)W3を作成し、単語W1と単語W3の認識結果を比較する。この際、単語W3の認識結果の単語評価値S3が単語W1に対する単語評価値S1よりも大きく、閾値以上と判断され、「NORTH YORK」が都市名と認識される。

【0075】

次に、本来複数の単語として切り出されるべきものが、1単語として切り出されることにより生じる誤認識の防止について説明する。

【0076】

すなわち、本来複数の単語として切り出されるべきものが、1単語として切り出されることにより、単語認識に失敗する場合がある。図9は、「TORONTO」「ON」と2単語に切り出されるべきものが1単語として切り出された例である。この場合オンタリオ州に「TORONTOON」という都市は存在しないため都市名認識に失敗する。

【0077】

このような単語の接触が発生している場合でも単語認識が行える住所単語認識処理の一例を図10のフローチャートに示す。住所認識制御部7から教えられた単語認識処理の開始位置から1単語ずつ住所単語辞書を使って認識していく。現在処理中の単語(オンタリオ州に続く都市名として「TORONTOON」)W1についてその単語W1がある基準を満たしているか否かを調べ、満たしている場合は単語W1を複数の単語(「TORONTO」)W2、単語(「ON」)W3に分割する。単語分割の基準としては、例えば単語を構成する各文字の間隔を

用いる。図 11 に示す例では、「TORONTO,」の直後が他と比べて文字間隔が大きくなっているため、その位置で単語を 2 つに分割している。例えば、垂直射影等により得られる単語ブロックにより文字間の距離が判別される。図 9 ～ 図 11 では説明を簡単にするため 2 単語を接続する場合のみを取り上げているが、3 単語以上に分割する場合もありうる。そして分割処理後にできた各単語に対して単語認識処理を行い、もっとも結果の良いものを選択する。

【0078】

そして、1 単語だけで単語認識した場合と、複数単語分割してから単語認識にかけた場合を比較し、結果の良かった方を採用する。認識結果の評価値があらかじめ設定してあるしきい値より低い場合はどちらの単語認識結果も採用せず、単語 W1 の次に書かれている単語を新たな単語 W1 として上記の処理を繰り返す。

【0079】

上記住所認識制御部 7 による住所単語認識処理について、図 10 に示すフローチャートを参照しつつ説明する。

【0080】

すなわち、住所認識制御部 7 は、住所単語認識処理を開始し、住所単語探索開始位置へ移動する (ST11)。たとえば、カナダ用の住所認識方法に設定されている場合、最終行の後ろから順に読んでいく。

【0081】

この際、住所認識制御部 7 は、認識処理にかけていない単語が存在しなかった場合 (ST12)、単語認識エラー処理へ移行する。

【0082】

上記住所認識制御部 7 は、ステップ 12 により認識処理にかけていない単語が存在した場合、単語を 1 つ選択し、選択した単語 W1 を与えられた地名辞書 (11、12、13) を用いて単語認識処理する (ST13)。たとえば、選択した単語 W1 が州名に対応する単語の場合、単語辞書 11 を用い、選択した単語 W1 が都市名に対応する単語の場合、上記州名に対応する単語辞書 12 を用い、選択した単語 W1 がストリート名に対応する単語の場合、上記都市名に対応する単語辞書 13 を用いる。

【 0 0 8 3 】

この結果、住所認識制御部 7 は、単語認識結果 A 1、単語評価値 S 1 とを算出する (S T 1 3) 。

【 0 0 8 4 】

次に、住所認識制御部 7 は、単語 W 1 が分割可能か否かを判断する (S T 1 4) 。

【 0 0 8 5 】

住所認識制御部 7 は、単語 W 1 が 2 つに分割可能と判断した場合、単語 W 1 を単語 W 2 と単語 W 3 を作成し (S T 1 5)、この作成した単語 W 2、W 3 を対応する地名辞書 (1 1、1 2、1 3) を用いて単語認識処理する (S T 1 6) 。

【 0 0 8 6 】

この結果、住所認識制御部 7 は、単語認識結果 A 3、単語評価値 S 3 とを算出する (S T 1 6) 。

【 0 0 8 7 】

これにより、住所認識制御部 7 は、単語 W 1 に対する一番高い単語評価値 S 1 と、単語 W 2、W 3 に対する単語評価値 S 3 が一番高い単語評価値 S 3 とを比較し、単語 W 2、W 3 に対する一番大きな単語評価値 S 3 が単語 W 1 に対する一番大きな単語評価値 S 1 よりも大きいか同じで、かつ単語 W 2、W 3 に対する一番大きな単語評価値 S 3 が所定のしきい値以上の際に (S T 1 7)、単語 W 2、W 3 に対する単語認識結果 A 3 を認識結果として出力する。

【 0 0 8 8 】

また、上記住所認識制御部 7 は、上記比較により単語 W 1 に対する一番大きな単語評価値 S 1 が単語 W 2、W 3 に対する一番大きな単語評価値 S 3 よりも大きく、かつ単語 W 1 に対する一番大きな単語評価値 S 1 が所定のしきい値以上の際に (S T 1 8)、単語 W 1 に対する単語認識結果 A 1 を認識結果として出力する。

【 0 0 8 9 】

また、上記住所認識制御部 7 は、上記ステップ 1 7、1 8 を満足しなかった場合、ステップ 1 2 に戻る。

【0090】

また、上記住所認識制御部 7 は、上記ステップ 14 において、単語 W1 が分割不可と判断した場合、単語 W3 に対する単語評価値 S3 を「0」とし（ST19）、ステップ 17 に進む。

【0091】

この場合の例を図 9 を用いて説明する。

【0092】

すなわち、単語（「TORONTOON」）W1 と、この単語 W1 を分割して単語（「TORONTO」）W2 と単語（「ON」）W3 を作成し、単語 W1 と単語 W2、W3 の認識結果を比較する。この際、単語 W2 の認識結果の単語評価値 S3 が単語 W1 に対する単語評価値 S1 よりも大きく、閾値以上と判断され、「TORONTO」がオンタリオ州に続く都市名として認識される。

【0093】

次に、単語絞り込み辞書のコンパクト化について説明する。

【0094】

すなわち、認識対象地域に存在する地名数が非常に多い場合、認識しようとする単語の文字認識結果の並びと地名单語辞書に登録されている地名单語との比較回数が増えて 1 単語当たりの単語認識時間が長くなる。この問題を解決する方法の 1 つとして、単語絞り込み辞書を用いて地名单語の数を減らす方法があることはすでに述べた。この単語絞り込み辞書は、上記単語辞書 4 あるいは住所認識制御部 7 に設けられる。

【0095】

この方式の難点は、認識対象地域内の全ての都市やストリート名について単語絞り込み辞書を用意した場合、単語絞り込み辞書の総容量が非常に大きくなることである。以下にこの問題を解決するための方法を説明する。

【0096】

例えば各都市ごとに、都市内に存在するストリート名の辞書を作成した場合、ストリート名辞書に登録される単語数は都市によって大きく異なる。図 12 に都市ごとのストリート数の一例を示す。このストリート数は、例えば上記都市名の

各辞書ごとに付与されている。

【0097】

ところで、単語絞り込み辞書を用いた単語候補の絞り込みは辞書に登録されている単語数が多い場合は有効であるが、単語数が少ない場合は意味がないだけでなく単語絞り込み処理に要する時間が無駄となるし、また単語絞り込み辞書そのものが不要である。例えば、単語絞り込み処理で得点の高い単語上位20位までを選択することにしていた場合、図12に示す都市A、Dは都市内に存在するストリート数が20未満なので絞り込みを実行しなくても探索パターン列と辞書パターン列の比較処理の回数は20未満で済む。

【0098】

上記単語辞書4に登録された単語数によって単語絞り込み処理を行うか否かを切替える処理の一例を図13のフローチャートに示す。

【0099】

すなわち、住所認識制御部7は、住所単語認識処理を開始し、認識対象とする地域および単語の種類によって単語辞書4を選択する(ST21)。ついで、住所認識制御部7は、選択した単語辞書4の登録単語数がしきい値T1(20)より多いか否かを判断する(ST22)。

【0100】

次に、住所認識制御部7は、登録単語数がしきい値T1より多いと判断した場合、単語絞り込み処理で評価値の高い辞書登録単語の上位T2位までを選択する(ST23)。

【0101】

ついで、住所認識制御部7は、単語絞り込み処理で選択した辞書単語と認識しようとする単語の比較処理を行う(ST24)。この結果、住所認識制御部7は、単語認識結果A、単語評価値Sとを算出する(ST24)。

【0102】

これにより、住所認識制御部7は、単語評価値Sが所定のしきい値S1以上の際に(ST25)、単語認識結果Aを認識結果として出力し、単語評価値Sが所定のしきい値S1以下の際に(ST25)、単語認識エラー処理となる。

【0103】

また、住所認識制御部 7 は、ステップ 22 において、登録単語数がしきい値 T1 より多いと判断した場合、単語辞書 4 に登録された全単語を選択する（ST26）。

【0104】

ついで、住所認識制御部 7 は、選択した辞書単語の全単語と認識しようとする単語の比較処理を行う（ST27）。この結果、住所認識制御部 7 は、単語認識結果 A、単語評価値 S とを算出する（ST27）。この後、住所認識制御部 7 は、ステップ 25 に進む。

【0105】

また、単語絞り込み辞書の総容量をできるだけ小さくするためには、登録単語数が少ない単語辞書用の絞り込み辞書を最初から用意しないことにすれば良い。

【0106】

そして、絞り込み辞書が存在する場合は絞り込み処理を行ってから単語認識処理を行い、絞り込み辞書が存在しない場合は絞り込み処理を行わずに単語認識処理を行うことにすれば良い。単語絞り込み辞書の有無によって単語絞り込み処理を行うか否かを切替える処理の一例を図 14 のフローチャートに示す。図 13 のフローチャートと同一部位には同一ステップを付与する。

【0107】

上記単語絞り込み辞書の有無によって単語絞り込み処理を行うか否かを切替える処理の一例を図 14 のフローチャートに示す。

【0108】

すなわち、住所認識制御部 7 は、住所単語認識処理を開始し、認識対象とする地域および単語の種類によって単語辞書 4 を選択する（ST21）。ついで、住所認識制御部 7 は、選択した単語辞書 4 用の単語絞り込み辞書が存在するか否かを判断する（ST22'）。

【0109】

次に、住所認識制御部 7 は、単語絞り込み辞書が存在すると判断した場合、単語絞り込み処理で評価値の高い辞書登録単語の上位 T1 位までを選択する（ST

23')。

【0110】

ついで、住所認識制御部7は、単語絞り込み処理で選択した辞書単語と認識しようとする単語の比較処理を行う(ST24)。この結果、住所認識制御部7は、単語認識結果A、単語評価値Sとを算出する(ST24)。

【0111】

これにより、住所認識制御部7は、単語評価値Sが所定のしきい値S1以上の際に(ST25)、単語認識結果Aを認識結果として出力し、単語評価値Sが所定のしきい値S1以下の際に(ST25)、単語認識エラー処理となる。

【0112】

また、住所認識制御部7は、ステップ22'において、選択した単語辞書4用の単語絞り込み辞書が存在しないと判断した場合、単語辞書4に登録された全単語を選択する(ST26)。

【0113】

ついで、住所認識制御部7は、選択した辞書単語の全単語と認識しようとする単語の比較処理を行う(ST27)。この結果、住所認識制御部7は、単語認識結果A、単語評価値Sとを算出する(ST27)。この後、住所認識制御部7は、ステップ25に進む。

【0114】

上記したように、国によって住所の記載書式が異なる場合でも、各国用専用の住所認識装置を設計すること無く、同一のハードウェアで構成するようにしたものである。

【0115】

これにより、ごくわずかな設定変更を行うだけで、世界各国の住所認識を行うことができる。

【0116】

【発明の効果】

以上詳述したように、この発明によれば、わずかな修正を加えるだけで各国の所在情報認識を実行できる所在情報認識方法と認識アプリケーションと記録媒体

と所在情報認識装置を提供できる。

【図面の簡単な説明】

【図 1】

この発明の実施形態の住所認識装置の概略構成を示すブロック図。

【図 2】

住所書式設定部の概略構成を示す図。

【図 3】

住所書式設定部の概略構成を示す図。

【図 4】

州名の単語辞書の例を示す図。

【図 5】

都市名の単語辞書の例を示す図。

【図 6】

ストリート名の単語辞書の例を示す図。

【図 7】

住所単語認識処理を説明するためのフローチャート。

【図 8】

住所単語認識処理において複数単語を接続して作成した単語を説明するための図。

【図 9】

住所単語認識処理において本来複数の単語として切り出されるべきものが、1 単語として切り出された例を説明するための図。

【図 10】

単語の接触が発生している場合でも単語認識が行える住所単語認識処理の一例を説明するためのフローチャート。

【図 11】

単語の分割を説明するための図。

【図 12】

都市ごとのストリート数の一例を示す図。

【図 1 3】

単語辞書に登録された単語数によって単語絞り込み処理を行うか否かを切替える処理の一例を説明するためのフローチャート。

【図 1 4】

単語絞り込み辞書の有無によって単語絞り込み処理を行うか否かを切替える処理の一例を説明するためのフローチャート。

【符号の説明】

S … 書状（紙葉類）

1 … 画像取り込み部

2 … 領域検出部

3 … 住所単語検出部

4 … 住所辞書

5 … 単語認識処理部

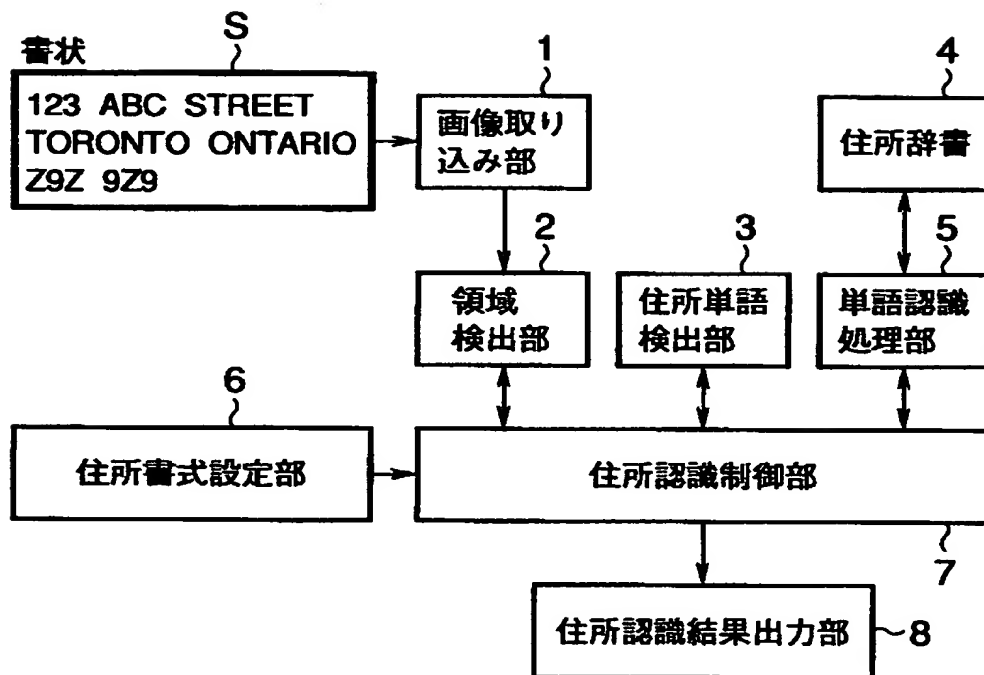
6 … 住所書式設定部

7 … 住所認識制御部

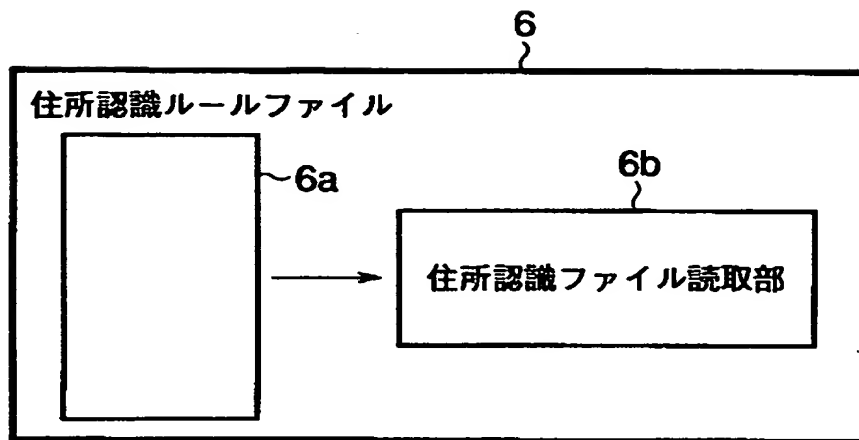
8 … 住所認識結果出力部

【書類名】 図面

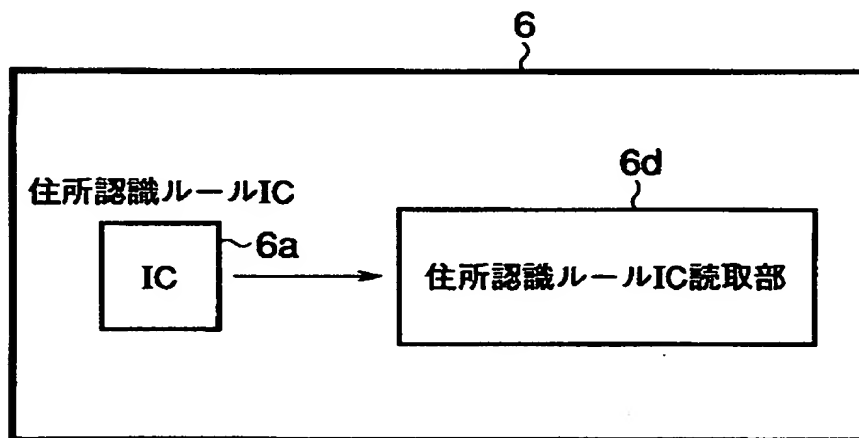
【図 1】



【図 2】



【図 3】



【図 4】

州名辞書 (カナダ)

1	ケベック	~11
2	オンタリオ	
3	サスカチワン	
4	アルバータ	
5	ブリティッシュコロンビア	
6	クロンダイク	

【図 5】

都市名辞書 (ケベック)

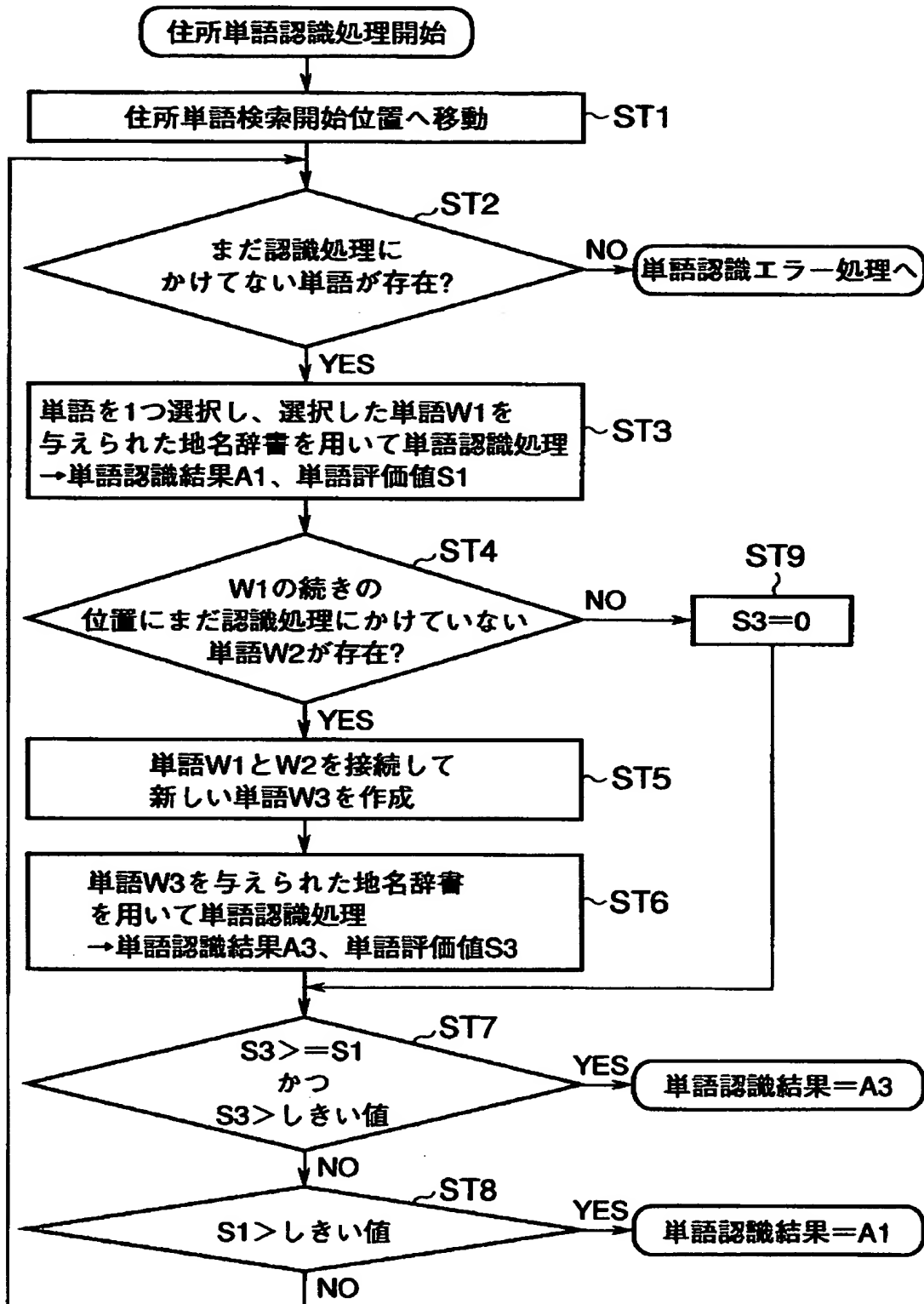
1	モントリオール	~12
2	オタワ	
3	コバルト	
4	キングストン	
5	ケベック	
6	ソーレル	

【図 6】

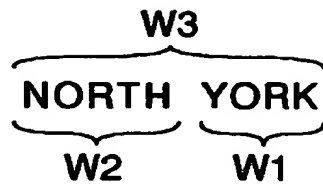
ストリート名辞書 (モントリオール)

1	サンバトリック	~13
2	オタワ	
3	ウィリアム	
4	キュイ	
5	サンジャック	
6	ノートルダム	
7	グレーヌ	

【図 7】



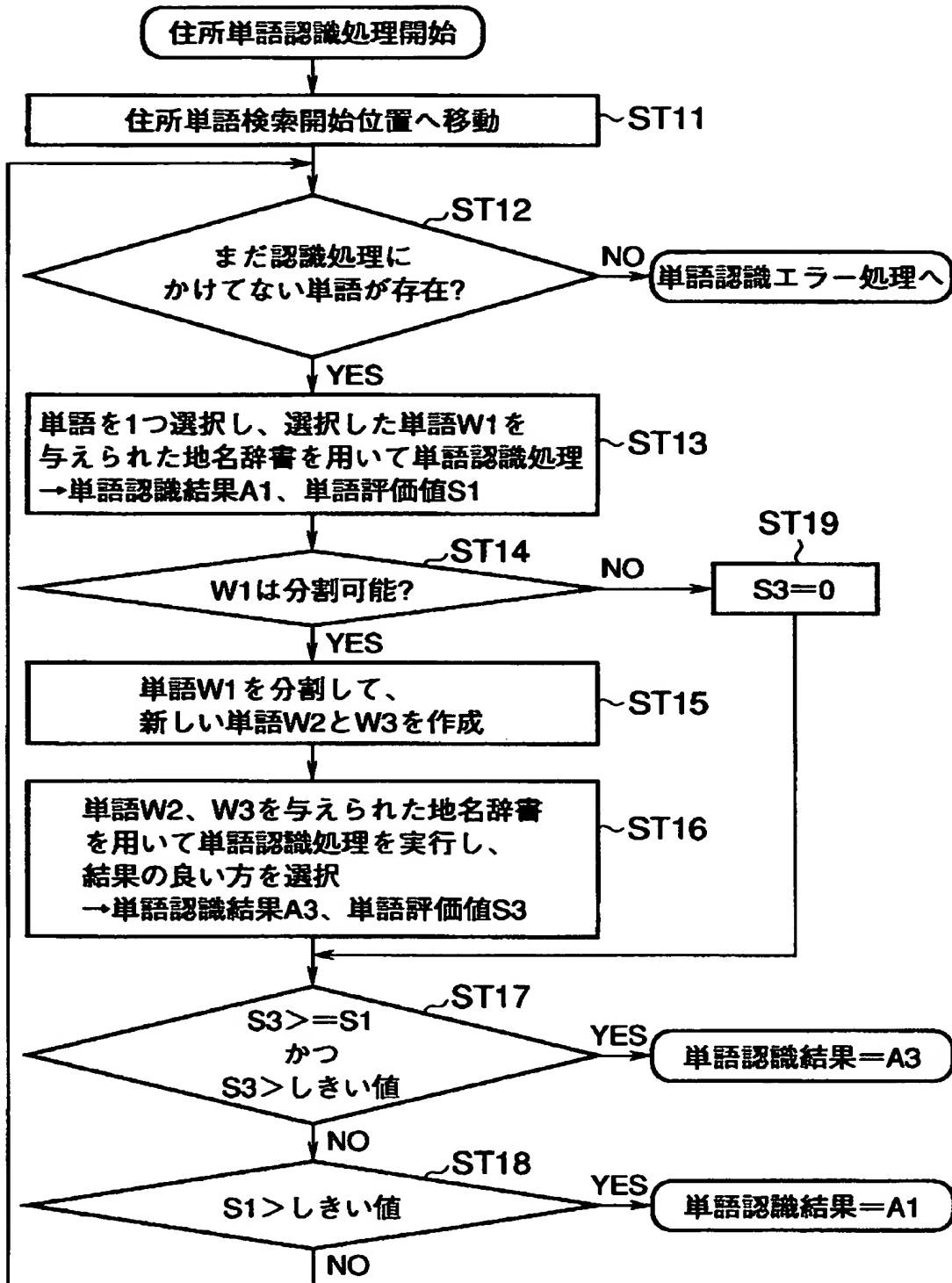
【図 8】



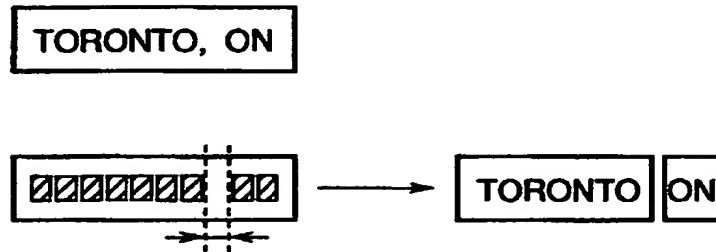
【図 9】

TORONTO, ON

【図 1 0】



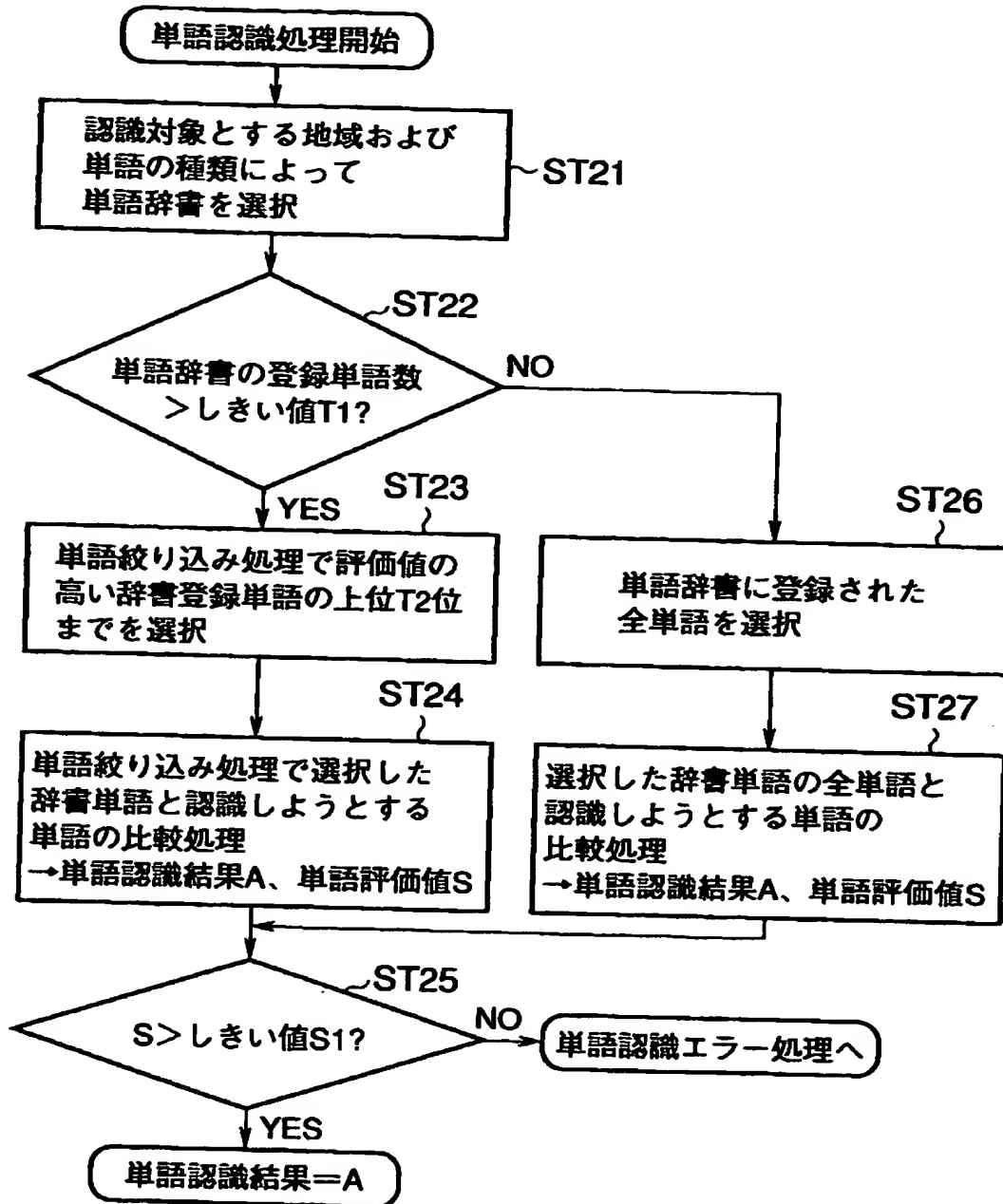
【図 1 1】



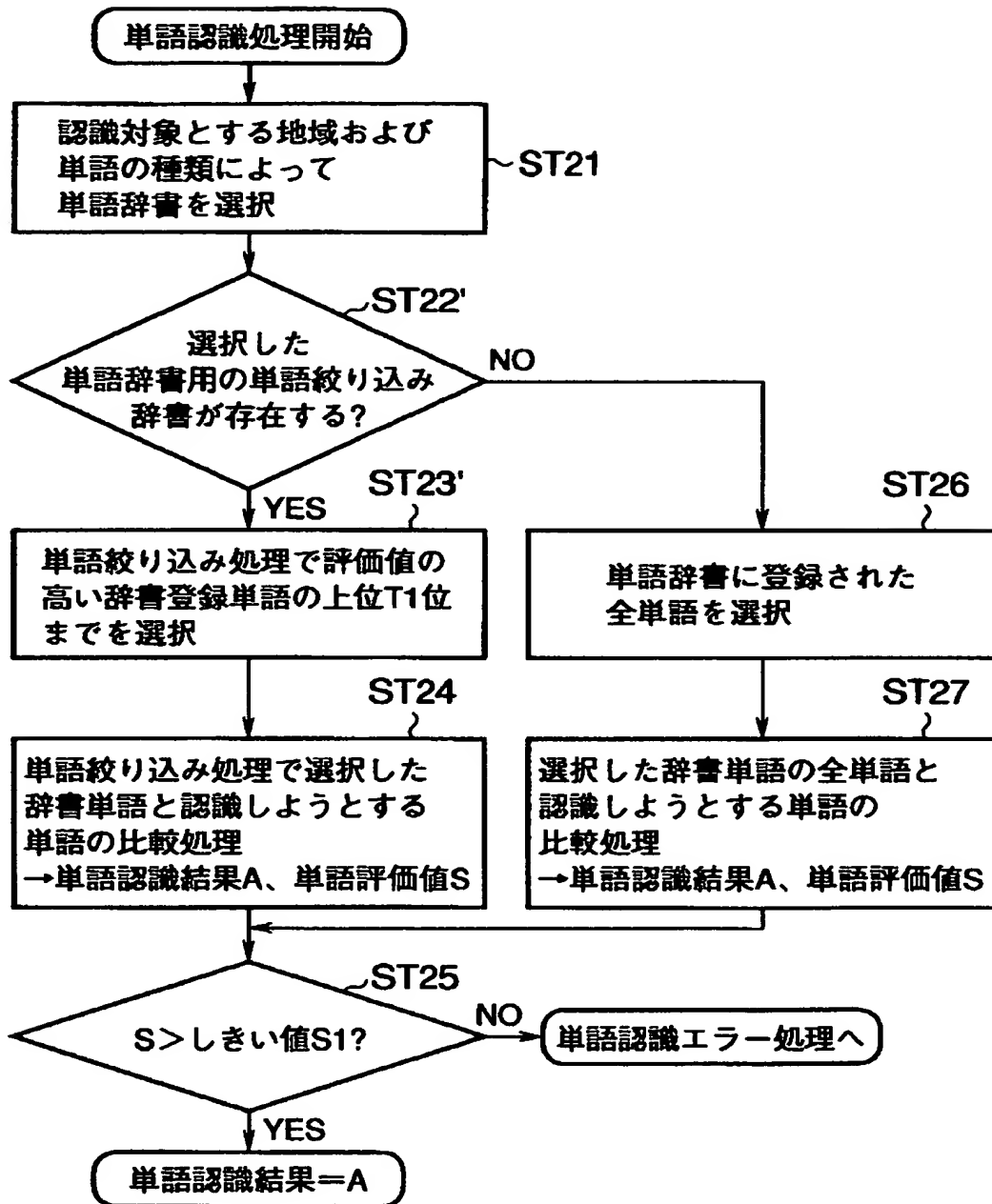
【図 1 2】

都市名	都市内のストリート数
A	10
B	1000
C	29
D	5

【図13】



【図 14】



【書類名】 要約書

【要約】

【課題】 この発明は、、わずかな修正を加えるだけで各国の所在情報認識を実行できる。

【解決手段】 この発明は、国によって住所の記載書式が異なる場合でも、各国用専用の住所認識装置を設計すること無く、同一のハードウェアで構成するようにしたものである。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000003078]

1. 変更年月日	1990年 8月22日
[変更理由]	新規登録
住 所	神奈川県川崎市幸区堀川町72番地
氏 名	株式会社東芝